DOCUMENT RESUME

ED 426 851 SE 061 204

AUTHOR	Ridgway, James
TITLE	From Barrier to Lever: Revising Roles for Assessment in Mathematics Education.
INSTITUTION	National Inst. for Science Education, Madison, WI.
SPONS AGENCY	National Science Foundation, Arlington, VA.
PUB DATE	1998-01-00
NOTE	9p.
CONTRACT	RED-9452971
AVAILABLE FROM	National Institute for Science Education, University of Wisconsin-Madison, 1025 W. Johnson Street, Madison, WI 53706; Web site: http://www.wcer.wisc.edu/nise
PUB TYPE	Collected Works - Serials (022) -- Reports - Descriptive (141)
JOURNAL CIT	NISE Brief; v2 n1 Jan 1998
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Academic Standards; *Educational Policy; Elementary Secondary Education; Evaluation Methods; Knowledge Base for Teaching; *Mathematics Education; Mathematics Teachers; *Science Education; *Student Evaluation
IDENTIFIERS	*Systemic Educational Reform; University of Wisconsin Madison

ABSTRACT

This brief presents the case for the reform of assessment systems as an essential component of systemic reform. The scope includes ways to share conceptual understanding among all stakeholders in the educational system, ways to evaluate the impact of systemic reforms in terms of student attainment of relevant goals, and ideas on how assessment can play a major role in guiding systemic reform. The formidable body of work done to develop new assessment systems is described. The ideas contained in this brief are generic and equally applicable to science education. (Author/DDR)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 426 851

NISE BRIEF

REPORTING ON ISSUES AND RESEARCH IN SCIENCE, MATHEMATICS, ENGINEERING, AND TECHNOLOGY EDUCATION

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

P. White

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

1

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

From Barrier to Lever: Revising Roles for Assessment in Mathematics Education

by James Ridgway

If one is to change classroom activities, it is important to align curriculum ambitions and assessment practices.

Educational reforms are likely to fail unless new forms of assessment are implemented that reflect new standards. Assessment systems play a key role in education: They provide the rewards for students, teachers and schools and have a major effect on what is taught and how it is taught. Appropriate assessment schemes can be powerful levers to support reform: assessment schemes that do not reflect new educational ambitions, however, are barriers to progress.

This Brief sets out the case for the reform of assessment systems as an essential component of systemic reform. The scope of this Brief includes ways to share conceptual understanding among all the stakeholders in the educational system, ways to evaluate the impact of systemic reforms in terms of student attainment of relevant goals, and ideas on how assessment can play a major role in guiding systemic reform. A great deal of work to develop new assessment systems has already been done in mathematics, some of which is described here. The

SE 061 204

NISE National Institute for Science Education

University of Wisconsin–Madison • National Center for Improving Science Education
Funded by the National Science Foundation

Vol. 2, No. 1 • January 1998

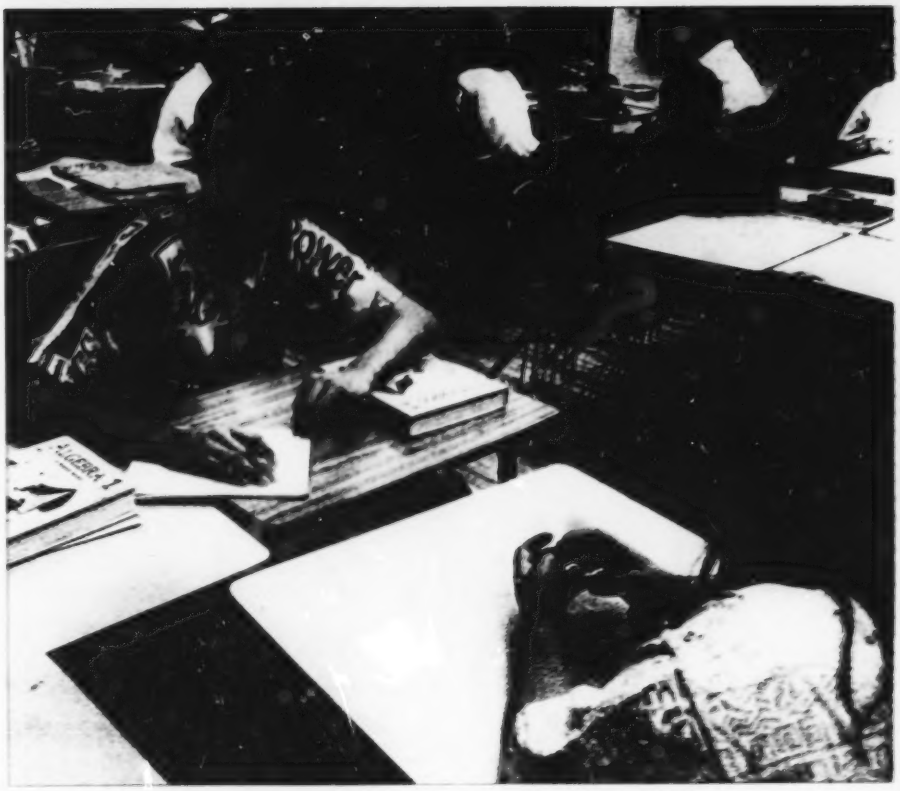
ideas set out in the Brief are generic and apply equally well to science education.

Why Do We Need to Revise Assessment Systems?

Public assessments establish the merit and competence of students, teachers and schools. Students and teachers are rewarded for good performance on the assessment system that is in place, no matter what its nature. Assessment, therefore, will drive (or constrain) educational activities (Messick, 1994; Ridgway & Passey, 1993). If one is to change classroom activities, it is important to align curriculum ambitions and assessment practices. New curriculum goals–such as making connections between mathematical topics and applying mathematics to situations–provide pressure to include less work in the curriculum on learning topics in isolation or on memorizing procedures and formulas without understanding.

However, since almost all "high stakes" assessments–such as state tests or SATs–assess mathematical technique on a narrow range of tasks, a serious conflict exists between new educational goals and current assessment practices. Most current high stakes assessment methods present a major barrier to educational reform. If reform goals are to be promoted, tests that measure decontextualized technical skills need to be replaced with tests that reflect the new intellectual agendas being initiated by professional societies, states and districts. As an issue of policy, the implementation of standards-based curricula should always be accompanied by the implementation of standards-based assessment.

A powerful lever for reform would be relevant feedback about the progress of Systemic Initiatives (SIs) and of new curricula. If an SI or a curriculum project sets out to promote fluent use of algebra in



Feedback about what is effective—and what is not effective—in promoting new educational goals is essential to progress.

realistic contexts, for example, there is little point assessing its success via a test of speeded arithmetic. However, even the tasks that comprise the National Assessment of Educational Progress (e.g., NCES, 1994) or the Third International Mathematics and Science Study (TIMSS) (e.g., Valverde & Schmidt) fall far short of what is required for the assessment of new curriculum goals.

The assessment vacuum poses a serious challenge to educational reform, because feedback about what is, and what is not, effective in promoting new educational goals is essential to progress. Establishing systems to assess educational progress will mean devoting considerable effort to the management, and reporting of information, both formally and informally.

What Is 'Balanced Assessment'?

Balanced assessment doesn't focus on a single theme, such as "technical skill" or "authentic performance," nor does it use a single method of assessment, such as multiple-choice tests or portfolios. This simple starting point opens up a debate at the heart of educational reform about the sorts of mathematics that students should acquire, and about how different aspects of performance should be recognized and rewarded. If assessment systems are to promote reform, they should be designed to exemplify the new curriculum goals. That is, the balance of assessment tasks should mirror the balance of the new curriculum.

One can identify a number of dimensions on which mathematics tasks differ, such as mathematical content (e.g., number, algebra, geometry); task type (e.g., technical exercise, nonroutine problem, creating a plan); or the circumstances of performance (e.g., multiple-choice item, 60-minute open-response task). Every task can be located within the space defined by these dimensions. Assessment is "balanced" if the assembly of tasks used to assess student performance samples each dimension in an appropriate manner.

Of course, the term "appropriate manner" brings with it another set of conceptual problems! There can be no unique description of individual tasks (for example, see the Skeleton Tower exercise described on page 5), nor of the domain of mathematics. Even if there were, it would not remain in place for long, as new branches of mathematics are invented, and as different areas of mathematics rise and fall in terms of their perceived relevance to school mathematics. Nevertheless, classifications are useful, because they draw attention to important aspects of mathematics that might otherwise be ignored. A number of classifications are available–from the National Council of Teachers of Mathematics, from the New Standards Project, from state and systemic initiatives, and from national curriculum framework documents around the world.

The National Science Foundation funded the Balanced Assessment project, based at the University of California at Berkeley, Harvard University, Michigan State University, and two universities in England, the University of Nottingham and the University of Lancaster. The project was aimed at producing ways to assess new curriculum goals. (Ridgway & Schoenfeld, 1994, details the project's rationale.) The "Dimensions of Balance" shown here are adapted from this project.

Dimensions of Balance

Mathematical Content will include some of the following:

Number and Quantity, including concepts and representation; computation, estimation and measurement; number theory and general number properties.

Algebra, Patterns and Function, including patterns and generalization; functional relationships (including ratio and proportion); graphical and tabular representation; symbolic representation; forming and solving relationships.

Geometry, Space and Shape.

Handling Data, Statistics and Probability.

Other Mathematics.

Mathematical Process, such as problem solving, reasoning and communication, will include some of the following:

Modeling and Formulating.

Transforming and Manipulating.

Inferring and Drawing Conclusions.

Checking and Evaluating.

Reporting.

Task Type will be one of the following:

Open Investigation.

Nonroutine Problem.

Design.

Plan.

Evaluation and Recommendation.

Review and Critique.

Re-presentation of Information.

Technical Exercise.

Definition of Concepts.

Goal Type will be one of the following:

Pure Mathematics.

Illustrative Application of Mathematics.

Applied Power over a Practical Situation.

Circumstances of Performance will include:

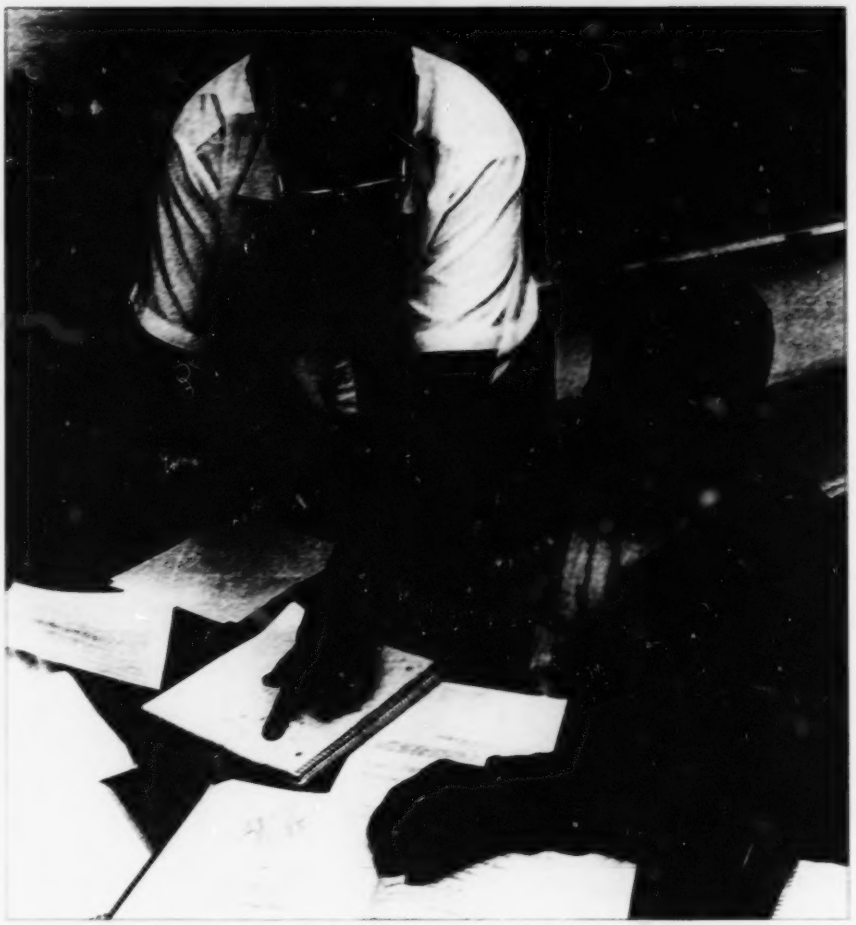
Task Length.

Modes of Presentation including, written, oral, video, computer.

Modes of Working, including individual, group, mixed.

Modes of Student Response, including written, built, spoken, programmed, performed.

Classification systems are not neutral descriptions. Rather, they are laden with beliefs about the nature of mathematics. An assessment system, through its tasks and scoring schemes, defines the nature of mathematics for its adoptive community. New frameworks and new assessment systems achieve wide acceptance only after a great deal of debate: They force an articulation of what is mathematically valuable, and indeed about the nature of mathematics. Vigorous debate may well be a good thing from the viewpoint of systemic reform, for it certainly shows that proposed changes are not just "business as usual, but with new badges."



The design and redesign of assessment systems that are aligned to reform initiatives can act as a "driver" of reform.

Developing New Tasks

New tasks require extensive supporting materials, such as:

A clear definition of the core mathematics. There is strong evidence that nonexperts judge tasks in terms of their surface features–"It's about playing with Lego," or "It's about drawing on T-shirts"–rather than in terms of their deep structures: generalization, proof, algebra, mathematical notation and communication. It is important to communicate the core mathematics to parents, to students and (sometimes) to teachers.

Administrative details. These details provide users with a way of quickly identifying candidate tasks. They include the student grade level for which it is designed; the mathematical background that students need in order to tackle the task; the length of time the task will take to administer; any materials that teachers need to assemble before they start; and the way students are to be grouped in order to perform on the task.

Examples of student work. These are essential to demonstrate the mathematics that students can produce and to show a variety of levels of performance.

A variety of scoring schemes. Some users (such as states and SIs) choose to employ holistic scoring schemes while

others prefer analytic schemes. It is important to respect these choices. Devising scoring schemes is a difficult process, which requires detailed analysis of student scripts and a good deal of thought. Users welcome exemplars of scoring schemes that fit their current practices. The guiding principle for the development of any scoring scheme is that procedures are in place to ensure that the scheme meets acceptable standards of reliability and validity. Evidence from studies in Vermont, for example, shows clearly that process skills can be assessed reliably, as do studies investigating the Connected Mathematics Project, which is described below. Issues surrounding the psychometrics of performance assessment are discussed in Phillips (1996).

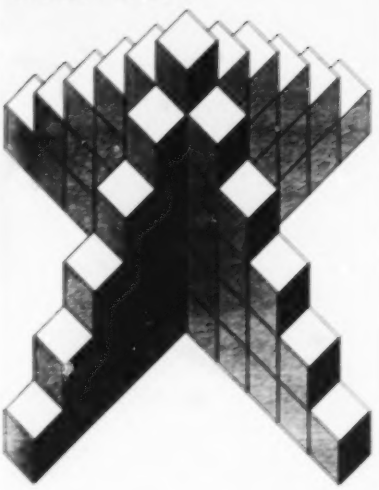
A "framework for balance" to guide users who want to assemble assessment instruments. This is an essential tool to help shape the local vision of mathematics and to plan the evolution of this vision over time.

What Do Tasks Assess?

Skeleton Tower, shown here, is an exercise designed for use in Grades 10 and 11. It presents a pattern and then calls for a generalization of the pattern, as well as explanation (hopefully a proof) of why the generalization holds true. Generalization, explanation and proof are deep mathematical ideas that give power over mathematical situations.

What does Skeleton Tower assess? That depends on how the student approaches the task. Every student who completes the task successfully must show abilities to generalize and to prove and to explain mathematical ideas. However, there are a number of different routes to success.

Skeleton Tower



1. How many cubes are needed to build this tower?
2. How many cubes are needed to build a tower like this, but 12 cubes high?
3. Explain how you worked out your answer to part 2.
4. How would you calculate the number of cubes needed for a tower *n* cubes high?

Some students answer the problem by considering mathematical series:

- They add up the blocks in each "wing" by counting (1+2+3+....), then multiply by 4 and add on the central column.
- Or they consider the tower as a set of horizontal slices, and count the blocks in each layer (1+5+9+....).

Some students answer the problem by rearranging the blocks in the Tower:

- They imagine tearing off two opposite "wings," which they turn upside down and stick onto the remaining structure to make a "wall."
- Or they imagine tearing the structure into four pieces and making a rectangle and a square with these pieces.

Students who use sums of series are showing their prowess in pure mathematics. Those who rearrange the structure, on the other hand, are showing their spatial skills. For students facing a task like this for the first time, solving the Skeleton Tower might involve a great deal of mathematical creativity. For students who have been exposed to a more open curriculum, the task may instead require the exercise of their process skills. And for still other students, Skeleton Tower may even be a memory task. Despite these differences, the mathematical demand of the task is unambiguous. Better answers can be distinguished from weaker ones, and student responses can be scored reliably.

Key Roles for Assessment in Education

Vivid Communication of New Goals

Tasks can provide clear illustrations of the educational goals of a reformed curriculum, and tasks with student work are often included in state framework documents and in documents for teachers, parents and students. Mathematics, however, is not a spectator sport: Vivid communication requires more than showing things for stakeholders to look at. It requires intellectual engagement with new ideas, such as working on tasks and scoring student work. An activity-based approach to communicating the goals of reform has been found to be successful when working with groups as diverse as mentor teachers, teachers, parents and community members.

The Balanced Assessment project includes the following activities. Sessions start with groups solving a problem and then sharing their solutions and the mathematical processes they used. This introduction gives everyone a good grasp of the mathematics contained in the task, and exposure to a range of solutions and explanations. Next, participants are given student scripts chosen to illustrate a range of

levels of performance. Individuals rank the scripts, and these ranks form the basis for discussions about the aspects of mathematics performance that are seen to be of value. After discussions, the next stage is to devise scoring schemes that use different methods, such as holistic judgement or allocating points for different parts. Working on scoring schemes and student scripts provides a sharp focus for discussions about the core mathematics and how student knowledge can be recognized and rewarded, which can lead participants to a deeper understanding of the educational ambitions underlying the reforms.

Customized Evaluation of Systemic Initiatives

The National Science Foundation emphasizes the need to reform assessment practices alongside changes in curricula and instructional practice. However, reform of state assessment mechanisms is often beyond the immediate reach of SIs (especially urban or rural SIs), and so most SIs are set in the context of traditional assessment systems. One approach to the design of assessment systems is to take account of existing tests (notably those mandated at state level) and to complement them with tasks designed to make up a fuller picture of students' mathematical competencies. This approach is being pursued in a pilot study in El Paso, Texas. Scores from the existing state test (the Texas Assessment of Academic Skills, or TAAS) will provide measures of mathematical technique; items from TIMSS and NAEP will provide a broader range of items for which national and international scores are available; and tasks from the Balanced Assessment task collection will provide a range of performance assessments.

An assessment system, through its tasks and scoring schemes, defines the nature of mathematics for its adoptive community.

Evaluating New Curricula

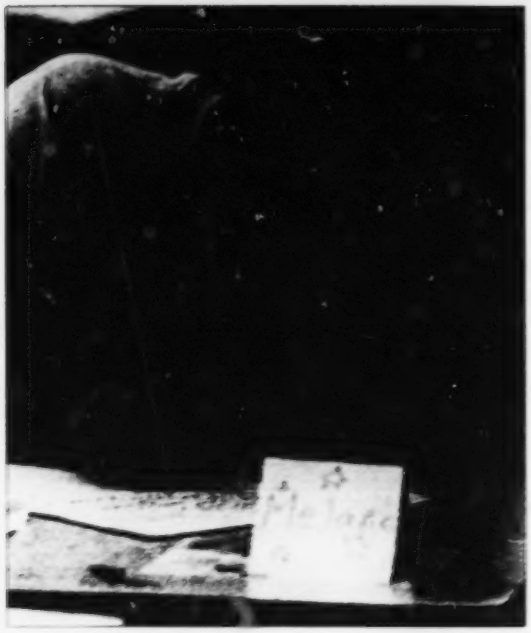
A broad and deep array of evidence is needed about what is and is not effective curriculum practice. However, attempts to assess the effectiveness of a new curriculum face a paradox: If traditional tests are used, they fail to measure the mathematical skills that the new curriculum is trying to teach; if tests are based on the curriculum materials, then participating students can hardly fail to outperform their rivals in the control group.

An approach that resolves this paradox has been used to evaluate the Connected Mathematics Project (CMP), which is funded by NSF to support middle school mathematics. An assessment scheme was produced that reflected broad NCTM ambitions for the middle school, rather than CMP ambitions themselves. Five parallel tests were assembled from the Balanced Assessment task bank for grades 6, 7 and 8, designed to assess student performance on tasks representing NCTM standards at those grade levels. The Iowa Tests of Basic Skills were used to assess technical skill.

Results showed that one year into the program, students following the CMP curriculum improved more than the control classes on the BA tasks, and there were no clear differences on the ITBS tests (Zawojewski, Hoover, & Ridgway, 1997). In the second and third years of the program, detailed analyses showed significant gains in ITBS scores for CMP students, compared with students in control classes.

The Design and Redesign of Assessment as a Driver of Systemic Reform

Lack of alignment between educational ambition and the assessment system will be a major hindrance to the reform process (Webb, 1997). Conversely, the design and redesign of assessment systems that are aligned to reform initiatives can act as a "driver" of reform. The revision of assessment methods over a period of time is likely to cause far less of a shock to an educational system than would the introduction of assessment instruments that match all the new educational goals at the outset. A policy of incremental change in assessment systems can promote change in



ways that allow professional development activities and curriculum development to keep up (Chrispeels, 1997). Collections of carefully validated tasks can allow assessment systems to be assembled that incorporate assessment practices already in place–such as state mandated tests or school-based portfolio work–so as to reflect short term goals on "balance." The risk of political backlash and large-scale resistance from educators, which might well overwhelm the efforts at reform, can be assessed and used to judge an acceptable pace for reform. The planned pace of reform is reflected (and made public) in the evolution of high-stakes assessment systems.

Conclusions

Exclusive dependence on standardized tests of technique in mathematics and science poses a substantial threat to educational reform. More balanced approaches to assessment have been developed and can be tailored to fit local ambitions and circumstances. Carefully constructed new styles of task are essential for communicating the intellectual heartland of new reforms to stakeholders and for monitoring—and sometimes for steering—the progress of reform.

❖❖❖

FOR FURTHER READING

Blank, R.K., Pechman, E. M., & Goldstein, D. (1996). *State mathematics and science standards, frameworks, and student assessments: What is the status of development in the 50 states?* Washington, DC: Council of Chief State School Officers.

Chrispeels, J. H. (1997). Educational policy implementation in a shifting political climate: The California experience. *American Educational Research Journal, 34*(3), 453-481

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(2), 13-23.

Phillips, G. W. (Ed.). (1996). *Technical issues in large scale performance assessment* (NCES 96-802). Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

National Center for Education Statistics. (1994). *NAEP 1994 trends in academic progress.* Washington DC: U.S. Department of Education.

The New Standards Performance Standards. (1997). (Volume 1, Elementary School; Volume 2, Middle School; Volume 3, High School). Pittsburgh: National Center on Education and the Economy and the University of Pittsburgh.

Ridgway, J., & Passey, D. (1993). An international view of mathematics assessment–through a class, darkly. In M. Niss (Ed.), *Investigations into assessment in mathematics education* (pp. 57-72). London: Kluwer.

Ridgway, J., & Schoenfeld, A. H. (1994). *Balanced assessment: Designing assessment schemes to promote desirable change in mathematics education.* Core paper for the E-mail Conference on Assessment, European Association for Research in Learning and Instruction.

Valverde, G. A., & Schmidt, W. H. (no date). *The Third International Mathematics and Science Study and the longitudinal study of school change: Issues and opportunities.* East Lansing: Michigan State University College of Education, U.S. National Research Center for TIMSS.

Webb, N. L. (1997). *Determining alignment of expectations and assessments in mathematics and Science education* (NISE Brief Vol 1, No. 2). Madison: University of Wisconsin-Madison, National Institute for Science Education.

Zawojewski, J. S., Hoover, M. N., & Ridgway, J. E. (1997). *Analysis of student performance on two assessment instruments in the Connected Mathematics Project.* Paper presented at the annual meeting of the American Educational Research Association, Chicago.

Very thoughtful reviews of earlier versions of this Brief were provided by Andrew Porter, NISE Co-Director; Norma Davila, University of Puerto Rico; William Firestone, Rutgers University; Kate Nolan, Milwaukee, Wisconsin; and Larry Suter, National Science Foundation.

BEST COPY AVAILABLE

Resources

TIMSS release items (about two-thirds of all the tasks they used) can be downloaded from: *http://www.csteep.bc.edu/timss*

Information on the NAEP 1996 Report Card can be obtained from the National Library of Education, Office of Educational Research and Improvement, U.S. Department of Education, 555 New Jersey Avenue, NW, Washington, DC 20208-5721

The NAEP Web site is *http://www.ed.gov/NCES/naep*

The NCTM Web site is *http://www.nctm.org*

Resources from the Balanced Assessment project, and consultancy on a range of issues related to the design of assessment methods, can be obtained from the NSF-funded Mathematics Assessment Resource Service at: *http://www.educ.msu.edu/MARS*

Shell Center. (1984). Problems with patterns and numbers. Manchester, England: Shell Center/JMB

James Ridgway is Director of the Mathematics Assessment Resources Service at Michigan State University. In 1996 and 1997 he was a Research Fellow at NISE. He teaches psychology at the University of Lancaster, England.

Photos by Jeff Miller

***NISE Brief* Staff**

Co-Directors	Andrew Porter
	Terrence Millar
Project Manager	Paula White
Editor	Leon Lynn
Editorial Consultant	Deborah Stewart
Graphic Designer	Rhonda Dix

This *Brief* was supported by a cooperative agreement between the National Science Foundation and the University of Wisconsin-Madison (Cooperative Agreement No. RED-9452971). At UW-Madison, the National Institute for Science Education is housed in the Wisconsin Center for Education Research and is a collaborative effort of the College of Agricultural and Life Sciences, the School of Education, the College of Engineering, and the College of Letters and Science. The collaborative effort also is joined by the National Center for Improving Science Education in Washington, DC. Any opinions, findings or conclusions herein are those of the author(s) and do not necessarily reflect the views of the supporting agencies.

No copyright is claimed on the contents of the *NISE Brief*. In reproducing articles, please use the following credit: "Reprinted with permission from the *NISE Brief*, published by the National Institute for Science Education, UW-Madison." If you reprint, please send a copy of the reprint to the NISE.

This publication is free on request. *NISE Briefs* are also available electronically at our World Wide Web site: http://www.wcer.wisc.edu/nise

National Institute for Science Education
University of Wisconsin-Madison
1025 W. Johnson Street
Madison, WI 53706
(608) 263-9250
(608) 263-1028
FAX: (608) 262-7428
Internet: niseinfo@macc.wisc.edu

Vol. 2, No. 1 January 1998

National Institute for Science Education
University of Wisconsin-Madison
1025 W. Johnson Street
Madison, WI 53706

Nonprofit Organization
U.S. Postage
PAID
Madison, Wisconsin
Permit No. 1622

DAVID HAURY
ERIC CLEARINGHOUSE FOR SCI.MATH &EE
1929 KENNY RD
COLUMBUS OH 43210-1080

BEST COPY AVAILABLE